DOCUMENT RESUME

ED 138 648 TM 006 307

AUTHOR Fishbein, Ronald L.; Shoemaker, David M.
TITLE Administrative and Political Ramifications of Multiple Matrix Sampling for the Practitioner.
PUB DATE [Apr 77]
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Item Sampling; *State Programs; *Testing Programs
IDENTIFIERS Michigan Educational Assessment Program; *Multiple Matrix Sampling

ABSTRACT

The administrative and political ramifications of a form of multiple matrix sampling where all students in the population are tested on a subtest is discussed in the context of a statewide testing program. The conclusion is drawn that large-scale testing programs will have to adopt a testing framework which samples items from the total item domain relevant to instruction to avoid contending constantly with political adversaries. Multiple matrix sampling is a tool which will foster the implementation of such a testing framework. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138648

ADMINISTRATIVE AND POLITICAL RAMIFICATIONS OF MULTIPLE MATRIX SAMPLING FOR THE PRACTITIONER*

Ronald L. Fishbein
David M. Shoemaker
United States Office of Education

Abstract

The administrative and political ramifications of a form of multiple matrix sampling where all students in the population are tested on a subtest is discussed in the context of a statewide testing program. The conclusion is drawn that large-scale testing programs will have to adopt a testing framework which samples items from the total item domain relevant to instruction to avoid contending constantly with political adversaries. Multiple matrix sampling is a tool which will foster the implementation of such a testing framework.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

*The opinions expressed herein are those of the authors. No official support or endorsement by the U.S. Office of Education is intended or should be inferred.

TM006 307

When evaluating the effectiveness of an instructional program, the primary emphasis is on assessing the performance level of a given group of students over an item domain which operationally defines those skills considered important. Here it is important to note that the psychometric procedure most appropriate for group assessment is distinctly different from that for evaluating the performance of each individual student. When the goal is to assess the achievement level of a group of students, a procedure more efficient than administering one test to all students is to divide the item domain into multiple subtests and administer each subtest to a subgroup of students selected randomly from the total group -- a procedure referred to typically as multiple matrix sampling.

The practical and statistical advantages of multiple matrix sampling have been documented by numerous authors (Shoemaker, 1973; Sirotnik, 1975; Womer, 1973). However, even though the procedure is well documented and the necessary computer programs are available, many potential practitioners continue to avoid the use of matrix sampling -- thinking incorrectly that the procedure is far too complicated to be implemented in practice. Our comments here will demonstrate that a form of multiple matrix sampling -- where each student in the population is tested on a subtest -- may be implemented by practitioners having various degrees of statistical expertise.

By example, we demonstrate here that practitioners have implemented this form of multiple matrix sampling with resulting administrative and data analysis advantages. Most important, however, is that the effective use of the technique may result in positive political ramifications which, ultimately, may have more of an effect on funding levels that the psychometric qualities of the program.

One Example of Implementation

Michigan Educational Assessment Program

The 1975-76 Michigan Educational Assessment Program (MEAP) used matrix sampling to gather statewide item and subtest information on recently developed fourth- and seventh-grade reading, mathematics, and health education items. These subtests were administered in addition to the every-pupil portion of the assessment program. The procedures followed by Michigan are an excellent example of the use of matrix sampling by practitioners.

The MEAP staff constructed 42 subtests (booklets) for both fourth- and seventh-grades. Each subtest consisted of approximately 20 items and took 30 minutes to administer. A single set of test administration instructions was developed for all subtests and made part of the regular Examiner's Manual at each grade level. The MEAP Scoring Service (Westinghouse Learning Corporation) printed and collated the subtests. For each grade, packages of 42 tests containing forms 1 through 42, followed by forms 1 through 42 were constructed. Packing instructions specified that the first school to be tested was to receive the required number of booklets beginning with form 1 from the first package and ending with whatever form number and package was appropriate. For example, if the first school

required 127 fourth-grade booklets, the first three packages would be used with the first booklet from the fourth package. If the next school required 75 booklets, then the first booklet in its shipment would start with form 2 and end with form 34. Except for the first school to be tested, this method provided a random starting point for each school. Additionally, each school testing coordinator was instructed to use the same procedure when distributing tests to teachers; that is, the first fourth-grade teacher in a school received the required number of booklets beginning with the first booklet in the fourth-grade shipment, with all form numbers having approximately an equal probability of being represented. Teachers were instructed to distribute the tests by walking up and down the rows of students giving the first booklet in the stack to the first student, the second booklet to the second student, etc., until all students had received a test booklet.

Administrative Advantages

For Districts, Schools, and Students. Districts and schools reacted very positively to this method because all students in all schools participated. District testing coordinators did not have to be concerned that only certain schools or students were to participate.[1] School administrators also were positive because all schools in the state containing fourth- and seventh-grade students were sharing equally in

[1] The 1974-75 MEAP used a one-stage stratified cluster sampling of schools for each tryout booklet. District testing coordinators had to be certain that the schools that were chosen to participate were given the correct booklets. There were several instances where a school administered the wrong form, where the wrong school was tested, or where a school was not tested at all because it had closed. The latter problem occurred because Michigan has many small rural districts where schools open and close from year to year and small districts sometimes consolidate with larger ones.

the testing burden. They did not have to contend with angry parents wanting to know why their school or certain students in their school had been chosen for "extra" testing. In fact, very few parents, students or teachers knew that the test section administered in separate booklets was for tryout purposes.[2]

For The Scoring Service And The MEAP. The staff of Westinghouse Learning Corporation found the procedure extremely easy to implement because all schools and students in the State containing fourth- and/or seventh-grade students participated. Clerical errors were reduced considerably and, to avoid shortages, all shipments contained approximately 10% more booklets than was indicated on the school's packing list. There was a cost-saving to Westinghouse, which was passed on to Michigan, because supervisory personnel were rarely needed for this aspect of the program and the services of a sampling statistician were not required. In addition, the problem of choosing substitute schools for those that could not test was avoided (see footnote 1).

Administrative Disadvantages

The chief administrative disadvantage of this procedure was that students were required to transfer their test booklet number onto their answer sheet, with the teacher checking this procedure for accuracy. Even though students were instructed to write their name on their test booklet, a small percentage of booklets were unscorable. This problem could be eliminated in a testing framework where matrix sampling is used exclusively. Answer sheets could contain the number of the test booklet

[2] In previous years, some sections of the battery were administered in separate booklets because these instruments were copyrighted tests bought from publishers.

with which they should be used and could be inserted into the test booklet by the scoring service before shipping to schools.

Data Analysis Advantages

One advantage of using the form of multiple matrix sampling where all students are tested, or any other random sampling procedure, is that the students responding to each test booklet may be treated legitimately as a simple random sample from the population of students. Because of this, commonly-used item and test analysis procedures may be employed and weighted data is unnecessary.[3]

Political Ramifications

In recent years, teacher organizations have been extremely critical of testing programs, particularly those using nationally standardized tests generating summary reports for each individual student. The crux of the problem is that when every student takes the same set of items, only a small portion of the relevant item domain can be tested. This is a legitimate concern for both norm-referenced and criterion-referenced testing programs.

The use of multiple matrix sampling is the most efficient method currently available to test the entire item domain for a given content area. If a testing program used this procedure, all legitimate groups could agree hopefully upon the domain of content because many subdomains for a given content domain could be tested. Furthermore, even though

[3] Proportional sampling is self-weighting and is useful in many situations. However, in a State like Michigan with approximately 600 independent school districts, where schools open and close without consulting the State education department, it is very difficult to maintain a proportional sample.

student achievement over a larger item domain may be assessed, the testing time per student would be well within tolerable limits. This would certainly help to quell the criticism that students and teachers are asked to spend too much time testing.

In addition to being a more viable alternative to large-scale assessment than testing all students over all items, or testing a sample of students over a sample of items, matrix sampling may be the only approach viable politically for testing programs to use given the power of teacher and school administrator organizations. Such groups will resist attempts to judge the effectiveness of their programs by a test consisting of items measuring only a small portion of that which should be measured -- only a small portion of the relevant item domain.

Although we have used the MEAP as an example of practitioners implementing matrix sampling procedures, Michigan did not test entire content domains. Michigan was used because of familiarity with the program. The California Assessment Program, however, applied the same basic technique as Michigan, but attempted to measure all reasonable aspects of the subject content domains tested. In California, there was little opposition to the program by teacher and administrator groups. This may be explained partially by the fact that educator groups helped select the many content subdomains that were tested and additionally reviewed all test items; also, testing time was drastically shortened and results were summarized at the state level only.

Although the focus of the discussion has been on the use of matrix sampling by practitioners for group assessment, we are in favor of individual assessment when appropriate. In our opinion, large-scale

testing programs should concentrate on gathering group information; individual assessment should be done at the local level on aspects of the content domain particularly relevant to local curricula. Individual achievement testing is most appropriately used within an instructional program to (1) determine whether a student has the prerequisite skills to begin an instructional program, and if he does, to place him within the program; (2) test the student on specific content subdomains; and (3) posttest the student over the entire instructional program. Hopefully, localities would have free access to items developed in their State.

Summary and Conclusions

Multiple matrix sampling is a technique which can be used effectively by practitioners to assess the achievement level of groups of students. The administrative and political ramifications of a form of the technique where all students in the population are tested on a subtest was discussed in the context of a statewide testing program. The conclusion was drawn that large-scale testing programs will have to adopt a testing framework which samples items from the total item domain relevant to instruction to avoid contending constantly with political adversaries. Multiple matrix sampling is a tool which will foster the implementation of such a testing framework.

References

Shoemaker, D.M. Principles and Procedures of Multiple Matrix Sampling. Cambridge, Mass.: Bollinger Publishing Company, 1973.

Sirotnik, K.A. Introduction to Matrix Sampling for the Practitioner. In Popham, W.J. Evaluation in Education: Current Applications. Berkeley, Ca.: McCutchan Publishing Corporation, 1975.

Womer, F.B. Developing a Large Scale Assessment Program. Denver, Colorado: Cooperative Accountability Project, 1973.